

**NATIONAL INSTITUTE on ALCOHOL ABUSE and
ALCOHOLISM (NIAAA)**

**Source and Accuracy Statement for Wave 1 of the 2001-2002 National
Epidemiologic Survey on Alcohol and Related Conditions**

Cite as:

Grant, B.F., Kaplan K., Shepard J., Moore T. Source and Accuracy Statement for Wave 1 of the 2001-2002 National Epidemiologic Survey on Alcohol and Related Conditions. National Institute on Alcohol Abuse and Alcoholism: Bethesda MD; 2003.

Source and Accuracy Statement for the 2001-2002 National Epidemiologic Survey on Alcohol and Related Conditions

Table of Contents

SOURCE OF DATA

Introduction	1
Purpose	1
Population and Frame	2
Sample Design	2
First Stage: PSU Selection	3
Second Stage: Within-PSU Selection	4
Third Stage	4
NESARC Final Sample Size	5
Estimation Procedure	6
Base Weight	6
Type A Household Noninterview Adjustment	6
Within-Household Sampling Adjustment	7
Usual Residence Elsewhere Adjustment	7
Type A Person Noninterview Adjustment	7
First Stage Adjustment	7
Second Stage Adjustment	7

ACCURACY OF THE ESTIMATES

Sampling Error	8
Nonsampling Error	8
Nonresponse and Survey Response Rates	8
Missing Data	10
Undercoverage	13
Comparability of Data	14
A Nonsampling Error Warning	15

Tables and Appendices

Table 1: Distribution of Collapsed NESARC Sample PSUs	4
Table 2: Impact of Oversampling on Variances	5
Table 3: Distribution of NESARC Sample Persons	7
Table 4: NESARC Response Rates	9
Table 5: NESARC Household Noninterview Adjustment Cells	9
Table 6: NESARC Person Noninterview Adjustments Cells	10
Table 7: Imputation of NESARC Variables	10
Table 8: Imputation Rates and Donor Cell Characteristics for NESARC Variables	11
Table 9: NESARC Coverage Ratios	14
Appendix A: Raking Cells	16
Appendix B: Glossary of Terms	19

SOURCE OF DATA

Introduction

The National Epidemiologic Survey on Alcohol and Related Conditions (NESARC) was designed to be a longitudinal survey with its first wave fielded in 2001-2002. The second wave is planned for 2004-2005 using the same respondents. The U.S. Bureau of the Census conducted the field work for this survey for the National Institute on Alcohol Abuse and Alcoholism (NIAAA), an agency of the National Institutes of Health. Data collection for Wave 1 began in August, 2001 and was completed by May, 2002 via computer assisted personal interviewing (CAPI).

Purpose

The major purposes of the NESARC are the following:

- to determine the magnitude of alcohol use disorders and their associated disabilities in the general population;
- to estimate the magnitude of health disparities and identify the determinants among subgroups of the population defined by gender, race/ethnicity, disability, age, low income, and socioeconomic status;
- to estimate the size, characteristics, and time trends of populations of special concern, including alcohol abusers and persons in the general population otherwise impaired by the use of alcohol;
- to provide more complete recording and tabulation of alcohol use disorders and their associated disabilities in important subgroups of the population;
- to estimate changes over time in alcohol use disorders and their associated disabilities and to identify factors that impact on their remission, chronicity, stability, and initiation;
- to increase our understanding of the natural history of alcohol use disorders and their associated disabilities;
- to determine the number and demographic characteristics of individuals seeking and receiving treatment through alcohol treatment programs and services, including those not otherwise represented in periodic surveys of treatment facilities or populations in treatment;
- to measure the number of people in the population in need of but not currently receiving treatment for alcohol use disorders and their associated disabilities;

- to provide information concerning barriers to alcohol related treatment services, particularly among low income groups, women, young adults, and race/ethnic minorities;
- to determine the economic impact of alcohol use disorders and their associated disabilities on impaired productivity in the workplace;
- to determine the magnitude and extent of binge drinking among college-aged young adults and to identify their characteristics and those risk factors influencing the initiation and remission of this hazardous drinking pattern;
- to determine the boundaries between safe and hazardous drinking levels and patterns for various types of alcohol use disorders and their associated medical, social, and psychological sequelae;
- to determine the associations between alcohol use disorders and their major physical and mental disabilities;
- to measure disability as a separate and distinct dimension of treatment need; and
- to determine the extent of major alcohol related mental and physical disabilities that are substance-induced disorders and differentiate those substance-induced disorders from those reflecting true, independent mental conditions.

Population and Frame

The NESARC is a representative sample of the United States population, including citizens and noncitizens. The target population of the NESARC is the civilian noninstitutionalized population, 18 years and older, residing in the United States and the District of Columbia (including Alaska and Hawaii). This includes persons living in households and the following noninstitutional group quarters: boarding houses, rooming houses, nontransient hotels and motels, shelters, facilities for housing workers, college quarters, and group homes.

The sampling frame of the NESARC sample for housing units (HUs) is the Census 2000/2001 Supplementary Survey (C2SS), a national survey of approximately 78,300 households per month conducted in 2000 and 2001 by the Bureau of the Census. The NESARC also included a group quarters (GQ) frame. The sampling frame for group quarters derives from the Census 2000 Group Quarters Inventory.

Sample Design

In order to understand the NESARC sample, it is important to understand the design of the C2SS. For the most part, C2SS primary sampling unit (PSU) definitions mirror the county-based PSUs used in the Census Bureau's Current Population Survey (CPS), with some changes made to account for changes in Metropolitan Statistical Area (MSA) and county definitions. The

approximately 2,000 C2SS PSUs in the universe consist of all 3,142 counties and county-equivalents in the United States. All PSUs with a 1996 population of 250,000 or more were designated as self-representing (SR). The C2SS also designated 42 additional counties as SR which were included for comparison with the Census 2000 results. All other PSUs were designated as nonself-representing (NSR) and stratified within state by the following variables:

- percent change in total PSU population between 1990 and 1996;
- number of vacant housing units (HUs) in 1990;
- percent change in number of HUs in PSUs between 1980 and 1990;
- number of renter occupied HUs in 1990;
- rural farm population in 1990;
- number of related children under 18 years below the poverty level in 1993;
- number of persons aged 10 to 19 years in 1990 who were not enrolled in school and were not high-school graduates;
- total Hispanic or Latino population in 1990 (in states where Hispanics or Latinos made up more than 10 percent of the projected total population for 2000): AZ, CA, CO, FL, IL, NV, NJ, NM, NY, TX; and
- total Black population in 1990 (in states where Blacks made up more than 10 percent of the projected total population for 2000): AL, AR, CT, DE, DC, FL, GA, IL, LA, MD, MI, MS, MO, NJ, NY, NC, OH, PA, SC, TN, TX, VA.

First Stage: PSU Selection. All SR PSUs were selected with certainty. For the NSR sample, two PSUs were selected per stratum with probability proportional to the size of the estimated 1996 population of the stratum. The resulting C2SS sample included 401 SR and 254 NSR, for a total of 655 PSUs. Note that for NESARC, sample was included from each of the 655 PSUs. However, to prevent potential respondent disclosure, some PSUs were collapsed so that the final NESARC data file shows 435 PSUs, 305 being SR and 130 being NSR.

Table 1 shows the distribution of the collapsed NESARC SR and NSR sample PSUs by each of the four geographic regions.

TABLE 1: DISTRIBUTION OF COLLAPSED NESARC SAMPLE PSUs

REGION	SR PSUs	NSR PSUs	Total
Northeast	66	6	72
Midwest	66	34	100
South	118	60	178
West	55	30	85
TOTAL U.S.	305	130	435

Second Stage: Within-PSU Selection. The primary source of the C2SS sample is the Census Bureau's Master Address File (MAF). For the C2SS unit frame sample, eligible housing units were extracted from the MAF within each sample PSU and a systematic sample of MAFIDs was chosen within each county. In some Census 2000 enumeration areas, an area frame sample was included. In about 66 counties, the C2SS included blocks where address listing and sampling procedures were performed.

For NESARC, information on race and ethnicity collected in the November 2000 - March 2001 C2SS panels was used to sort housing units within each PSU into three groups: Hispanic¹, non-Hispanic Black, and Other (non-Black, non-Hispanics). A housing unit was defined as Hispanic if at least one Hispanic person resided there. If there were no Hispanics and at least one Black resident, the housing unit was defined as non-Hispanic Black. All remaining housing units were placed in the Other category. Non-Hispanic Black and Hispanic housing units were selected at higher rates than Other housing units. Group quarters units were converted to HU equivalents and sampled together with the HUs. Each "hit" determines the HU or GQ to be interviewed for the NESARC.

Third Stage. For each household, one sample person was selected randomly from a roster of persons living in the household. For GQs, interviewers obtained a list or roster of persons residing in the group quarters and selected persons based upon the position of their name on the list. In households where young adults resided (age 18 to 24 years), the young adults were sampled at a rate of 2.25 times that of other members of the household. This rate was determined by calculating the expected change in variances for different rates. Table 2 illustrates the expected effects of various rates of oversampling young adults on variances of the estimates. The proportions in the table represent the ratio of the variance resulting from sampling young adults differentially compared to a uniform sampling rate overall.

¹ Hispanics may be of any race.

The oversampling rate that was selected could yield a considerable decrease (25 percent) in the variance of estimates for young adults, while the offsetting increases in the total population variance (4 percent) and for the variance for adults 25 years of age and older (5 percent) were acceptable.

TABLE 2: IMPACT OF OVERSAMPLING ON VARIANCES

OVERSAMPLING RATE	CHANGE IN VARIANCES		
	18-24	25+	Total
1.25	0.915	1.014	1.003
1.50	0.855	1.026	1.009
1.75	0.810	1.036	1.017
2.00	0.776	1.045	1.027
2.25	0.749	1.053	1.037
2.50	0.727	1.060	1.049
2.75	0.709	1.065	1.060
3.00	0.693	1.071	1.072

Oversampling of young adults is critical in the NESARC in order to understand the nature of heavy drinking patterns and adverse consequences of drinking (e.g., motor vehicle crashes, binge drinking) among the young adult population with a view towards developing targeted intervention and prevention programs. Oversampling of non-Hispanic Blacks and Hispanics in the NESARC is equally important in providing: (1) accurate and precise estimates of major survey variables; (2) adequate numbers for reliable statistical analysis; and (3) appropriate representation of each major race/ethnic subgroup of the United States population.

NESARC Final Sample Size

Overall, the NESARC sample resulted in a total of 43,093 completed interviews. The oversampling procedure for non-Hispanic Black households increased the number of non-Hispanic Black households from about 12.3 percent to 19.1 percent (8,245 cases). Oversampling of Hispanic households increased the number of Hispanic households from 12.5 percent to 19.3 percent (8,308 cases). Table 4 illustrates the distribution of NESARC sample person interviews by age, race, ethnicity, and sex categories.

TABLE 3: DISTRIBUTION OF NESARC SAMPLE PERSONS

AGE	HISPANIC		NON-HISPANIC BLACK		OTHER		ALL PERSONS		
	M	F	M	F	M	F	M	F	Total
18-24 years	730	754	395	584	1,285	1,451	2,410	2,789	5,199
25-29 years	412	521	217	460	778	1,079	1,407	2,060	3,467
30-34 years	473	634	294	523	1,067	1,301	1,834	2,458	4,292
35-39 years	473	628	320	584	1,196	1,449	1,989	2,661	4,650
40-44 years	385	468	317	543	1,332	1,395	2,034	2,406	4,440
45-49 years	333	355	336	469	1,207	1,318	1,876	2,142	4,018
50-54 years	248	297	304	465	1,051	1,242	1,603	2,004	3,607
55-59 years	162	219	227	360	847	1,032	1,236	1,611	2,847
60-64 years	149	172	159	325	717	846	1,025	1,343	2,368
65 + years	357	538	472	891	2,275	3,672	3,104	5,101	8,205
18 + years	3,722	4,586	3,041	5,204	11,755	14,785	18,518	24,575	43,093

Estimation Procedure

The NESARC weights sample results to form national level estimates. The final weight is the product of the NESARC baseweight and other individual weighting factors.

Base Weight. The base weight is the inverse of the probability of selection of a sample housing unit or housing unit equivalent for GQs. For NESARC this is the product of the conditional probabilities of selection at the first two stages of sampling. These two stages consist of the selection of the PSU within a stratum (for NSR PSUs) and the housing unit within the sample PSU.

Type A Household Noninterview Adjustment. The Type A household noninterview adjustment factor accounts for those households in which no household information was obtained, no sample person was selected, and no interview was obtained from the household. This type of noninterview can result from a refusal, the absence of all household members, or an inability to locate the household. This factor is computed by grouping PSUs into noninterview cells by MSA status, race/ethnicity, and region (see Table 5) and dividing the weighted number of eligible households by the weighted number of interviews. This factor inflates the weight of

interviewed households in the cell to account for these noninterviewed households.

Within-Household Sampling Adjustment. The within-household sampling adjustment factor accounts for the selection of one sample person from a household as well as the oversampling of young adults. In households which contain both young adults and adults 25 years and older, the probability of selection for each young adult is 2.25 times what the probability of selection is for an adult ages 25 or over. The factor is defined as the inverse of the probability of selection of the sample person in the household.

Usual Residence Elsewhere (URE) Adjustment. The URE adjustment accounts for errors where the first person listed was sometimes mistakenly labeled as a URE. UREs are not eligible for interviews. For household with two or more eligible persons, the factor is computed within 8 cells defined by region and sex. This factor inflates the weight of the sampled person to account for ALL persons in the household including the person who was counted as a URE. For single eligible person households, the weights of the sample persons were inflated to account for the URE person households because no one in the household was selected for sample.

Type A Person Noninterview Adjustment. The person noninterview adjustment factor accounts for those households in which a sample person was selected, but no interview was obtained. This type of noninterview can result from a refusal by the sample person, the absence of the sample person, etc. The factor is computed within 12 cells defined by the four geographic regions and three types of households - all young adults (ages 18 through 24), all adults 25 years and older, or both young adults and adults 25 years and older (see Table 6). This factor inflates the weight of interviewed persons in the cell to account for these noninterviewed persons.

First-Stage Ratio Adjustment. This adjustment affects only sample persons within NSR PSUs. The purpose of the first-stage ratio adjustment is to reduce the component of the variance arising from selecting two PSUs to represent an entire stratum. This factor was computed within 16 cells defined by the four geographic regions, race (Black/non-Black status), and MSA/non-MSA status.

Second-Stage Ratio Adjustment. The second-stage factor adjusts weighted sample results to agree with independent estimates of the civilian noninstitutional population of the United States by region, age, sex, race, and ethnicity. The adjusted estimate is called the post-stratification ratio estimate. The independent estimates are calculated based on information from the 2000 Decennial Census of Population and Housing; statistics on births, deaths, immigration, and emigration; and statistics on the size of the armed forces. While the adjustment mainly helps to reduce the component of bias resulting from sampling frame undercoverage, it also frequently reduces sampling variance.

In this adjustment, the weights are adjusted to agree with two sets of controls (see Appendix A). Adjusting the weights to match one set of controls can cause differences between the estimates and the other controls, so an iterative procedure, known as raking, is used to match the weights to agree with both sets of controls simultaneously. In NESARC the weights are adjusted to match

the first set of controls (region, race, age, and sex), then the second set (region, ethnicity, age, and sex). This process is repeated until the estimates converge to agree with both sets of controls.

ACCURACY OF THE ESTIMATES

A sample survey estimate has two types of error: sampling and nonsampling. The accuracy of an estimate depends on both types of error. The nature of the sampling error is known given the survey design. The full extent of the nonsampling error, however, is unknown.

Sampling Error

Since the NESARC estimates come from a sample, they may differ from figures from a complete census using the same questionnaires, instructions, and enumerators. This possible variation in the estimates due to sampling error is known as “sampling variability.” Standard errors are primarily measures of sampling variability, although they may include some nonsampling error. For the NESARC, standard errors may be computed using standard statistical software packages that account for the design effects of complex sample surveys (e.g., SUDAAN and WesVAR).

Nonsampling Error

All other sources of error in the survey estimates are collectively called nonsampling error. Nonsampling errors can be attributed to many sources including the following:

- The inability to obtain information about all cases in the sample (nonresponse).
- Differences in the interpretation of questions.
- Respondent inability or unwillingness to provide correct information.
- Respondent inability to recall information.
- Errors made in data collection, such as in recording or coding data.
- Errors made in processing the data.
- Errors made in estimating values for missing data.
- Failure to represent all units with the sample (undercoverage).

Three types of nonsampling error in NESARC have been examined to a limited extent: nonresponse, missing data, and undercoverage.

Nonresponse and Survey Response Rates

The effect of nonresponse cannot be measured directly, but one indication of its potential effect is the nonresponse rate. Survey response rates ($1 - \text{nonresponse rate}$) are commonly used in assessing the potential for nonresponse error.

The NESARC overall survey response rate consists of three parts. First is the NESARC household response rate (89%). Household nonresponse occurred when no interview was

obtained from the household and a sample person was never selected. The second part of the NESARC overall survey response rate is the person response rate (93%). Person nonresponse occurred when a sample person was selected but never was interviewed. NESARC estimates were adjusted at the household and person level to account for nonresponse from refusals, absences, and unlocated housing units. Tables 5 and 6 show the characteristics used for these two adjustments.

The third part of the NESARC overall survey response rate is the sample frame response rate (99%), reflecting the initial C2SS nonresponse rate (1%). This nonresponse rate included nonresponders from the C2SS who were eligible to be interviewed in NESARC but were not. The sample frame nonresponse rate reflects the nonresponse from the five panels, or months, of the C2SS used to draw the NESARC sample.

The overall survey response rate in NESARC is derived by multiplying the NESARC household response rate (89%) by the NESARC person response rate (93%) and the C2SS response rate (99%) as shown in Table 4. The overall survey response rate for the NESARC was 81%.

TABLE 4: NESARC RESPONSE RATES

Household Response Rate	89%
Person Response Rate	93%
Sample Frame Response Rate	99%
NESARC Overall Survey Response Rate	81%

TABLE 5: NESARC HOUSEHOLD NONINTERVIEW ADJUSTMENT CELLS

MSA STATUS/Race Stratum		REGION			
		Northeast	Midwest	South	West
MSA	Hispanic				
	Black				
	Other				
NON-MSA	Hispanic				
	Black				
	Other				

TABLE 6: NESARC PERSON NONINTERVIEW ADJUSTMENT CELLS

HOUSEHOLD TYPE	REGION			
	Northeast	Midwest	South	West
All Young Adults				
All Adults 25+				
Mixed Household				

Missing Data

The potential impact of residual item nonresponse in the NESARC, like most other surveys, is corrected through the use of imputation. Imputation rates are computed as the ratio of the number of eligible people who had a value imputed for that item to the number of people eligible to respond to that item. The process by which values for missing or inconsistent data are determined is complex and varies by survey. In general, there is a continuum of certainty about the probable content of a missing data item. Analysts are confident about assigning values to a missing data item when related information is available on the same person record. For example, first name may be used to assign a value of sex. Such imputations are often known as assignments since they do not rely on data from a separate record. Confidence is lower when values for missing or inconsistent items cannot be derived from the same person record and must come from other responding households believed to have similar characteristics. When such donors are used, the item is said to have been allocated. Such imputation is considered less accurate than using information derived from the housing unit or person record. In the NESARC, age, sex, and marital status variables were imputed using both assignment and allocation. Table 7 shows the imputation rates for these variables. All other variables imputed in the NESARC used only the allocation method. The imputation rates along with characteristics constituting the donor cells for these variables are shown in Table 8.

TABLE 7: IMPUTATION OF NESARC VARIABLES¹

VARIABLE	IMPUTATION RATE
SEX	0.14%
AGE	1.12%
MARITAL STATUS	0.59%

¹ This computation excludes item nonresponse that was addressed through assignment methods.

**TABLE 8: IMPUTATION RATES AND DONOR CELL CHARACTERISTICS
FOR NESARC VARIABLES**

VARIABLE	IMPUTATION RATE	CHARACTERISTICS
S1Q1D1: Race Indian	1.43%	Other Indian in household S1Q1E: Ethnicity (Indian, Western Hemisphere, Hispanic, other)
S1Q1D2: Race Asian	1.43%	Other Asian in household S1Q1E: Ethnicity (Asian, Pacific Islander, Middle East, other)
S1Q1D3: Race African-American	1.43%	Other African-Americans in household S1Q1E: Ethnicity (African-American or American or non-Spanish Caribbean; American Indian; Spanish Caribbean; other)
S1Q1D4: Race Native Hawaiian or Pacific Islander	1.43%	Other Hawaiian or Pacific Islander in household S1Q1E: Ethnicity (Guam or Samoan or Pacific Islander; Filipino or Indonesian or Malaysian; other)
S1Q1D5: Race White	1.43%	Other White in household S1Q1E: Ethnicity (American Indian, Asian, African-American, Pacific Islander, Middle Eastern, other)
S1Q1C: Hispanic Origin	0.33%	S1Q1D5: Race White S1Q1D3: Race African-American S1Q1E: Ethnicity (value = 8, 9, 11, 35, 36, 39, 43, 44, 53, 58; other)
S1Q16A: Highest Grade	1.09%	If S1Q17D has valid answer then S1Q17D (grade level) alone; Otherwise use: SEX AGE (8 groups) S1Q1D3: African-American S1Q1C: Hispanic Origin
S1Q7A1-S1Q7A14: Present situation (Imputed as a group)	0.47%	S1Q8A: Work in the last 12 months SEX AGE (4 groups) S1Q1D3: African-American S1Q1C: Hispanic Origin
S1Q7B: Full-time/part-time student	0.50%	AGE Employment status (full-time, part-time, other)
S1Q7C: Where lived while in school	4.22%	S1Q7B: full-time/part-time student SEX MARITAL: married, not married AGE (18, 19, 20, 21, 22, 23+)

VARIABLE	IMPUTATION RATE	CHARACTERISTICS
S1Q7D: Grade level during 2000-2001 school year	4.56%	S1Q7B: full-time/part-time student S1Q7C: live while in school MARITAL: married, not married AGE (8 groups)
S1Q8A: Work in the last 12 months	0.13%	AGE (8 groups) SEX
S1Q8B: Ever worked	0.09%	AGE (7 groups) SEX
S1Q9A: Industry of current or most recent job	2.75%	SEX S1Q6A: Highest grade (4 groups)
S1Q9B: Occupation of current or most recent job	2.63%	SEX S1Q1D3: Race African-American S1Q6A: Highest grade (4 groups)
S1Q9C: Type of employer of current or most recent job	2.46%	S1Q9A: Industry
S1Q10A: Personal income	11.11%	AGE SEX S1Q6A Employment status (full-time, part-time, other) S1Q9B
S1Q11A: Family income	10.11%	S1Q10A SEX Adult relatives in the household
S1Q12A: Household income	2.04%	S1Q11A AGE Adult non-relatives in the household
S1Q131: Receive Social Security in last 12 months	0.88%	AGE S1Q8B: Ever worked S1Q3B: Ever married
S1Q132: Receive SSI in last 12 months	0.87%	S1Q7A8: (on disability) S1Q171, S1Q172, S1Q181, S1Q182
S1Q133: Receive TANF in last 12 months	0.79%	S1Q12B: Household income SEX S1Q1D3: Race African-American Children in the household
S1Q134: Receive WIC in last 12 months	0.70%	S1Q12B: Household income SEX S13Q7B: Pregnancy in the last year AGE (18-24, 25-64, 65+) Children in the household

VARIABLE	IMPUTATION RATE	CHARACTERISTICS
S1Q14A: Receive food stamps in last 12 months	0.78%	S1Q12B: Household income SEX AGE (18-24, 25-64, 65+) Children in the household
S1Q14B: Amount of food stamps received in last 12 months	6.07%	S1Q12B: Household income SEX AGE (18-24, 25-64, 65+) Children in the household
S1Q14C1: Currently covered by Medicare	0.76%	AGE (18-61, 62-64, 65+)
S1Q14C2: Currently covered by Medicaid	0.87%	S1Q12B: Household income S1Q132: Receive SSI S1Q133: Receive TANF
S1Q14C3: Currently covered by CHAMPUS	0.74%	SEX AGE S1Q1D3: Race African-American In military or most recent job in military (S1Q9A = 14 or S1Q9B = 14 or S1Q9C = 6)
S1Q114C4: Currently covered by other medical insurance	0.85%	Covered by other insurance (S1Q14C1, S1Q14C2, S1Q14C3) Employment status (full-time, retired, part-time, other)

Undercoverage

The concept of coverage in the survey sampling process is the extent to which the total population that could be selected for sample “covers” the survey’s target population. NESARC undercoverage results from missed housing units and missed persons within interviewed households. Compared to the level of the updated estimates from the 2000 Decennial Census, NESARC undercoverage is about 13.1%.

Undercoverage varies with age, sex, and race/ethnicity. Generally, undercoverage is larger for males than for females and larger for Blacks than for Others.

As described in the estimation section, the weighting procedure uses ratio estimation whereby sample estimates are adjusted to independent estimates of national population by age, race, sex, and Hispanic ancestry. This weighting adjustment partially corrects for the bias due to undercoverage, but biases may still be present when people who are missed by the survey differ from those interviewed in ways other than age, race, sex, and Hispanic ancestry. How this weighting procedure affects other variables in the survey is not precisely known. All of these considerations affect comparisons across different surveys or data sources.

A common measure of survey coverage is the coverage ratio, the estimated population before post-stratification divided by the independent population control. A coverage ratio of 1.0 means that NESARC coverage is equivalent to the coverage of the adjusted 2000 Decennial Census figures. A ratio of .90 means that NESARC coverage is 10 percent worse. A coverage ratio greater than 1.0 indicates NESARC overcoverage of that group. Table 7 shows NESARC coverage ratios for age, sex, race, and ethnicity groups. The figures in the table represent the ratio of the NESARC estimate to the updated 2000 Decennial Census estimate for that group.

TABLE 9: NESARC COVERAGE RATIOS

AGE	HISPANIC		BLACK		OTHER		ALL PERSONS		
	M	F	M	F	M	F	M	F	Total
18-24 years	0.864	0.861	0.629	0.742	0.798	0.819	0.794	0.815	0.804
25-29 years	0.736	0.866	0.636	0.815	0.709	0.888	0.705	0.874	0.790
30-34 years	0.736	0.929	0.716	0.813	0.780	0.886	0.766	0.883	0.825
35-39 years	0.803	0.999	0.665	0.842	0.800	0.922	0.785	0.921	0.855
40-44 years	0.811	0.892	0.631	0.883	0.866	0.897	0.835	0.895	0.865
45-49 years	0.956	0.912	0.808	0.840	0.885	0.907	0.884	0.899	0.891
50-54 years	0.833	0.941	0.857	0.940	0.815	0.933	0.821	0.934	0.879
55-59 years	0.774	0.847	0.921	1.064	0.848	0.901	0.849	0.914	0.882
60-64 years	0.924	0.861	0.761	1.081	0.917	0.888	0.904	0.906	0.905
65 + years	0.970	0.906	0.956	0.968	0.978	0.964	0.976	0.961	0.967
18 + years	0.823	0.903	0.744	0.874	0.847	0.906	0.834	0.902	0.869

Comparability of Data

Data obtained from the NESARC and other sources are not entirely comparable. This results from differences in interviewer training and experience and in differing survey processes. This is an example of nonsampling variability not reflected in the standard errors. Therefore, caution should be used when comparing results from different sources.

Nonsampling Error Warning

Since the full extent of the nonsampling error is unknown, one should be particularly careful when interpreting results based on small differences between estimates. Even a small amount of nonsampling error can cause a borderline difference to appear significant or not, thus distorting a seemingly valid hypothesis test. Caution should also be used when interpreting results based on a relatively small number of cases. Summary measures may not reveal useful information when computed on a base smaller than 145,000.

APPENDIX A: RAKING CELLS

STEP 1: by Region, Race, Age, Sex

MaleMidwest

Black	White	Other
18-29	18-24	18-39
	25-29	
30-34	30-34	
35-39	35-39	40+
40-44	40-44	
45-49	45-49	
50-54	50-54	
55-59	55-59	
60+	60-64	
	65+	

FemaleMidwest

Black	White	Other
18-24	18-24	18-29
25-29	25-29	
30-34	30-34	30-39
35-39	35-39	
40-44	40-44	40+
45-49	45-49	
50-54	50-54	
55-59	55-59	
60+	60-64	
	65+	

Northeast

Black	White	Other
18-24	18-24	18-29
25-29	25-29	
30-34	30-34	30-39
35-39	35-39	
40-49	40-44	40-49
	45-49	
50+	50-54	50+
	55-59	
	60-64	
	65+	

Northeast

Black	White	Other
18-24	18-24	18-29
25-29	25-29	
30-34	30-34	30-39
35-39	35-39	
40-49	40-44	40-49
	45-49	
50+	50-54	50+
	55-59	
	60-64	
	65+	

Male**South**

Black	White	Other
18-24	18-24	18-24
25-29	25-29	25-34
30-34	30-34	
35-39	35-39	35-49
40-44	40-44	
45-49	45-49	
50-59	50-54	50+
	55-59	
60+	60-64	
	65+	

Female**South**

Black	White	Other
18-24	18-24	18-24
25-29	25-29	25-34
30-34	30-34	
35-39	35-39	35-49
40-44	40-44	
45-49	45-49	
50-59	50-54	50+
	55-59	
60+	60-64	
	65+	

West

Black	White	Other
18-29	18-24	18-24
	25-29	25-29
30-39	30-34	30-34
	35-39	35-39
40-49	40-44	40-44
	45-49	45-49
50-59	50-54	50-59
	55-59	
60+	60-64	60+
	65+	

West

Black	White	Other
18-29	18-24	18-24
	25-29	25-29
30-39	30-34	30-34
	35-39	35-39
40-49	40-44	40-44
	45-49	45-49
50-59	50-54	50-54
	55-59	55-59
60+	60-64	60+
	65+	

RAKING STEP 2: by Region, Ethnicity, Age, Sex

The following cells are the same for both sexes.

Midwest

Hispanic	Non-Hispanic
18-24	18-24
25-29	25-29
30-34	30-34
35-39	35-39
40-49	40-44
	45-49
50+	50-54
	55-59
	60-64
	65+

South

Hispanic	Non-Hispanic
18-24	18-24
25-29	25-29
30-34	30-34
35-39	35-39
40-44	40-44
45-49	45-49
50-54	50-54
55-59	55-59
60-64	60-64
65+	65+

Northeast

Hispanic	Non-Hispanic
18-24	18-24
25-29	25-29
30-34	30-34
35-39	35-39
40-44	40-44
45-49	45-49
50-54	50-54
55-64	55-59
	60-64
65+	65+

West

Hispanic	Non-Hispanic
18-24	18-24
25-29	25-29
30-34	30-34
35-39	35-39
40-44	40-44
45-49	45-49
50-54	50-54
55-59	55-59
60-64	60-64
65+	65+

APPENDIX B: GLOSSARY OF TERMS

The following are definitions of terms that are used in this report.

Accuracy - The difference between the sample result and the true population value. Attributes are measured in terms of sources or error (e.g, coverage, sampling, nonresponse, response, measurement, and processing).

American Community Survey (ACS) - The proposed replacement for the decennial long form survey. When fully implemented, it will collect the detailed demographic data traditionally collected on the decennial census long form from 3 million houses a year, located in every county, American Indian and Native Alaskan area, and Hawaiian Homeland, as well as Puerto Rico. These data will provide updates on detailed characteristics about our nation every year, rather than only once every ten years. Implementation of the ACS will enable the 2010 census to collect only short form information. *See 2000 Supplementary Sample (C2SS).*

Allocation - Refers to imputation required when values for missing items cannot be derived from the existing response record. In these cases, the imputation must be based on other techniques such as using answers from other persons in the household, other responding households, or persons believed to have similar characteristics. Such donors are reflected in a table referred to as an allocation matrix. *See also Assignments and Imputation.*

Assignment - Imputation method in which values for a missing item or an inconsistent item can be derived from the sample housing unit or person record. For example, a first name is often used to determine the sex of a person.

Census 2000 Group Quarters Inventory - A file of all the GQs enumerated in Census 2000.

Computer Assisted Personal Interviewing (CAPI) - Method or mode of data collection consisting of the interviewer asking questions displayed on a portable computer screen and entering the answers directly into the computer.

Current Population Survey (CPS) - Monthly sample survey of the U.S. population that provides employment and unemployment figures as well as current data about other social and economic characteristics of the population. Collected for the Bureau of Labor Statistics by the Census Bureau.

Housing Unit - A house, apartment, mobile home, or trailer, group of rooms or single room occupied as a separate living quarters or if vacant, intended for occupancy as a separate living quarters. The definition of separate living quarters for Census 2000 is that the occupants live separately from any other individuals in the building and have direct access from outside the building or through a common hall. Additional criteria, such as the presence of a kitchen or cooking equipment for the exclusive use of the occupants, were used to define a housing unit in previous censuses.

Imputation - Process used to resolve problems of missing, invalid, or inconsistent response identified during editing. Responses or missing values on the edited record are changed to ensure that a plausible, internally coherent record is created. *See also Allocation and Assignment.*

Long Form - The decennial census questionnaire containing 100-percent (short form) and sample questions. Sent to a sample of addresses in the census, long forms typically contain the short-form person and housing items that all households are asked to provide. Whereas short-form items are generally limited to basic demographic and housing questions, long-form items cover such topics as income, employment, veteran status, transportation to work, education, and others.

Master Address File (MAF) - The Census Bureau's permanent list of addresses for individual living quarters that is linked to the TIGER data base. *See also Topologically Integrated Geographic Encoding and Referencing (TIGER) System.*

Measurement Error - Error when the response received differs from the "true" value due to the respondent, the interviewer, the questionnaire, the mode of collection, or the respondent's record-keeping system(s).

Metropolitan Statistical Area (MSA) - An area defined by the Office of Management and Budget as a federal statistical standard. An area qualifies for recognition as an MSA if it includes a city of at least 50,000 population or an urbanized area of at least 50,000 with a total metropolitan area population of at least 100,000.

Nonsampling Error - Generally means all error (e.g., coverage, nonresponse, interviewer, respondent, instrument, mode) that does not arise merely because a sample of the population is measured.

Nonself-Representing (NSR) PSU - A PSU that does not represent a whole stratum.

Overcoverage - Extent to which an estimate from the frame includes more elements than the target population, including duplicate elements.

Primary Sampling Unit (PSU) - A major statistical unit, generally consisting of a county or group of contiguous counties, used for the Census Bureau's sample surveys.

Sampling Frame - Any list of device that, for purposes of sampling, delimits, identifies, and allows access to the sampling units which contain elements of the sampled population. The frame may be a listing of persons, housing units businesses, records, land segments, etc. One sampling frame or a combination of frames may be used to cover the entire sample population.

Sampling Error - The difference between a sample result and the result from a complete count taken under the same conditions.

Self-Representing (SR) PSU - A PSU that is treated as a separate stratum and included in the sample, usually because it is highly populated.

Stratum, strata - A sampling stratum is a grouping or classification that has a similar set of characteristics based on the previous census.

2000 Supplementary Survey (C2SS) - A Census 2000 experiment demonstrating that the American Community Survey (ACS) can be implemented nationwide at the same time, but separate from, the decennial census. Conducted in 1,203 counties, it is the test vehicle for reporting on the operational and technical performance of the ACS. *See American Community Survey (ACS).*

Topologically Integrated Geographic Encoding and Referencing (TIGER) System - A computer database that contains all census-required map features and attributes for the United States and its possessions, plus the specifications, procedures, computer programs, and related input materials required to build and use it.

Undercoverage - The extent to which an estimate from the frame includes fewer than the sampled population.

Weighting - A series of survey adjustments. Survey data are traditionally weighted to adjust for the sample design, the effects of nonresponse, and to correct for survey undercoverage error. *See also Nonresponse and Undercoverage.*